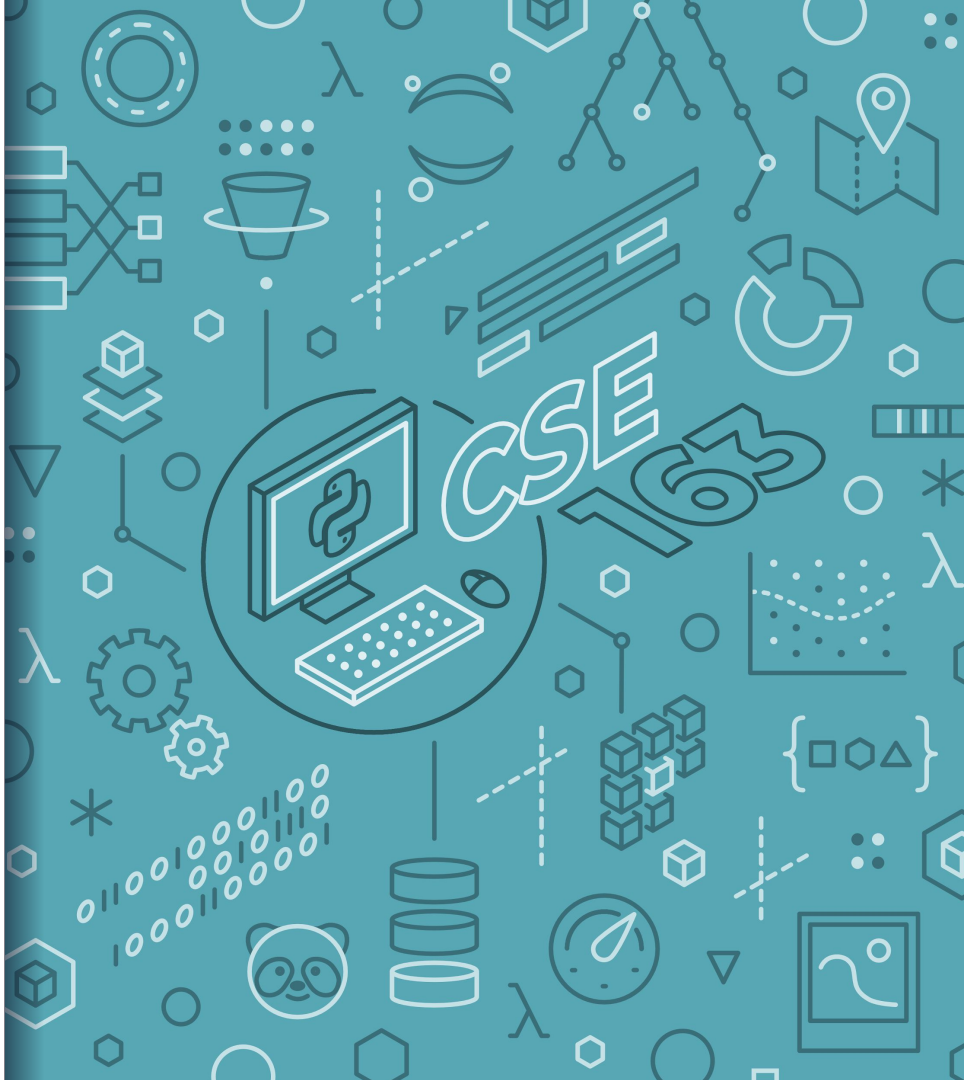
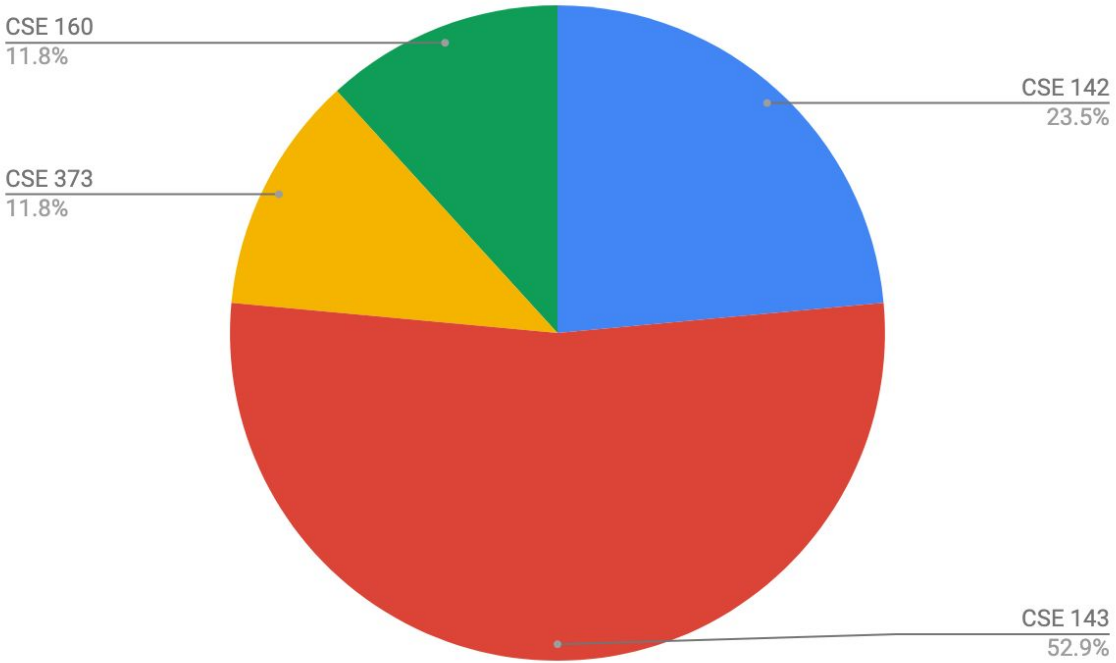
CSE 163

Victory Lap

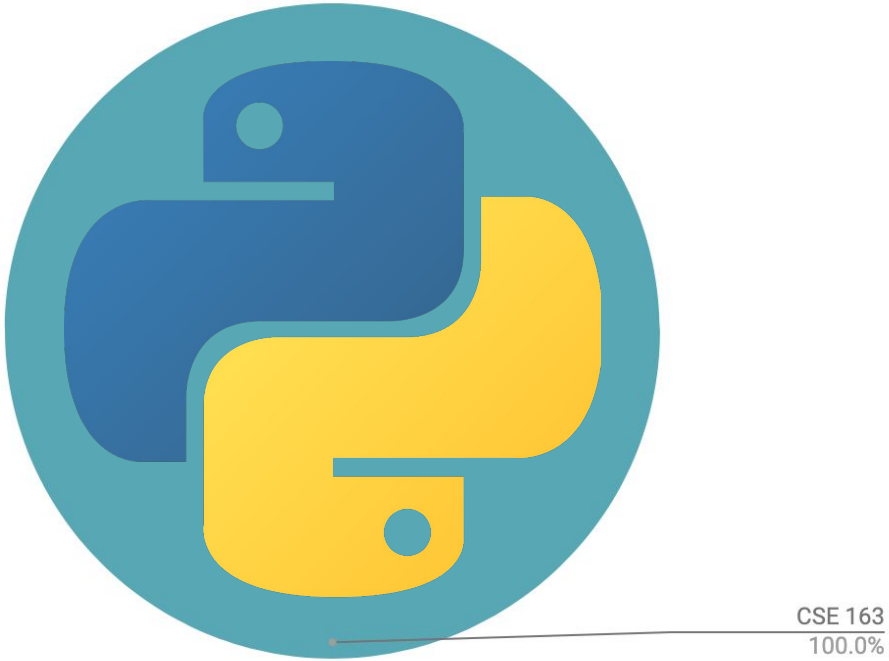
Hunter Schafer



Where were you from?



Where are you now?



Overview

What is this class?

Competencies

1. **More advanced programming concepts** than in CSE 142 or CSE 160 including how to write bigger programs with multiple classes and modules.

2. How to **work with different types of data**: tabular, text, images, geo-spatial, etc.

3. Ecosystem of **data science tools** including Jupyter Notebook and various **data science libraries** including scikit-image, scikit-learn, and pandas data frames.

4. Basic concepts related to code complexity, efficiency of different types of data structures, and memory management.

Competency 1

More advanced programming concepts than in CSE 142 or CSE 160 including how to write bigger programs with multiple classes and modules.

- Testing code
- Data structures: Lists, sets, dictionaries, tuples
- Classes and Objects
- Modules and Packages
- Anonymous functions (lambdas) and functional programming

Competency 2

How to **work with different types of data**: tabular, text, images, geo-spatial, etc.

- Tabular
- Time Series
- Unstructured Text
- Geo-spatial
- Images
- Webpages

Competency 3

Ecosystem of **data science tools** including Jupyter Notebook and various **data science libraries** including scikit-image, scikit-learn, and pandas data frames.

- Jupyter Notebooks
- Visual Studio Code + Anaconda
- Pandas
- Seaborn / Matplotlib
- Scikit-learn
- Geopandas
- Numpy
- Scikit-image
- Requests

Competency 4

Basic concepts related to code complexity, efficiency of different types of data structures, and memory management.

- Algorithmic efficiency
- Code profiling
- Computer memory
- Hashing
- Spatial Indices

- Why Python is slow and why that’s okay

“Extras”

These are things we learned that don’t necessarily fall under one of the categories lined above, but are necessary to support those goals

- Principles of data visualization
- Machine learning
- **Ethics**
- Joining data from multiple sources

What’s Next?

Future Classes

Web Programming: CSE 154

Machine Learning

- CSE/STAT 416 *
- STAT 435
- INFO 371

Societal Implications of Data Science

- SOC 225

Data Management

- CSE 414
- INFO 445

Data Visualization

- CSE 412
- INFO 474
- HCDE 411

Data Science Courses at UW

Online Resources

The internet is teeming with tutorials and online classes that teach topics in data science and data processing!

- Coursera
- Allen School Course Catalog
- Berkeley Data8 and DS100
- Towards Data Science (on Medium, hit or miss)

Projects

It's not really possible to list out all the things you can do with what you've learned in this class. I usually do a project to learn a new tool.

- Learn a new library!
  - Data Visualization: Bokeh or Altair
  - Natural Language Processing: NLTK or spaCy
  - Machine Learning: Tensorflow, Keras, PyTorch
  - Images: Open-CV
- Learn a new Language! Python is the most prominent language for data science, but it doesn't hurt to know more!
  - R - Numerical Processing
  - Scala - Compatible with Java, Nice syntax
  - Julia - New and up-and-coming language
  - Javascript - Language of the web
- Contribute to open source libraries!

You have already done a personal project this quarter!

Jobs

The field for jobs in computing is rapidly expanding

- A quickly growing subset of these jobs are ones that focus on processing large amounts of data
- Data Scientist is a relatively new term to describe these people

How to get a job in tech?

- Career fairs around campus
- Start looking early!
- Cast a broad net and don't be scared of getting turned down
- For those just starting out
  - Microsoft Explorer Program
  - Google Engineering Practicum

My Job Experience

- Redfin
  - Job: Full-stack engineer
  - Languages: Java + Javascript
- Socrata (Seattle City Data)
  - Job: Mostly data science, a little bit of search
  - Machine Learning: Python
  - Search Backend: Scala + ElasticSearch
- Sift Science
  - Job: Machine learning infrastructure
  - Language: Java + Python
  - Libraries: Spark

Research

You are all in the very unique experience of having some real marketable skills for doing undergrad research

- You have learned the language and the tools used by many researchers across the university
- Talk about what you know about processing data in Python, and you'll be surprised by the number of professors that NEED your help

Teaching

One of the best ways to become an expert in something is to teach it to someone else

We are looking to greatly expand the class next year so we will be looking for TAs for Winter/Spring quarter!

If you want me to reach out to you in Fall when I start looking for TAs, please fill out this interest form!

HUGE Thanks to the TAs

AA (12:30)

Erika Wolfe
eywolfe@cs

Josh Ervin
joshue@cs

AB (1:30)

Dylan Jergens
dylanj7@cs

Erik Hoberg
ehoberg@cs

AC (2:30)

Nicole Riley
nriley16@cs

Joely Nelson
joelyn@cs